UDC 551.577.36.551.509.54

A NOTE ON THE DERIVATION OF A SCALE MEASURE FOR PRECIPITATION EVENTS

CHARLES F. ROBERTS

Fire and Atmospheric Sciences Research, Forest Service, U.S. Department of Agriculture, Washington, D.C.

ABSTRACT

The importance of the relationship between scale and predictability suggests the need to define rational scale measures for various weather phenomena. A simple probability model is used to relate a characteristic space scale measure to point frequencies.

1. INTRODUCTION

In a previous study of the predictability of weather elements (Roberts 1969), the author found some degree of confirmation of a widely held hypothesis regarding the existence of a relationship between the space scale of weather phenomena and the accuracy with which the phenomena are predicted. A recent survey of National Weather Service forecasting performance also showed that the skill of the precipitation forecasts released to the public differed widely from one part of the country to another and between seasons and times of the day (Roberts et al. 1969). Since the basic ability of forecasters should be essentially the same for the entire country, these differences must be attributable to variations in some feature of the precipitation regimes characterizing different parts of the country, possibly involving both time and space scale factors. These findings suggest the need to develop a means of classifying and describing weather phenomena according to characteristic time and space scale measures.

In addition to shedding considerable light on the problems of predicting and describing weather phenomena, such information might have a wide variety of applications. Uses for precipitation scale data appear in such diverse fields as weather forecasting, hydrology, weather modification experiments, forest fire danger assessment, and forest fire control, just to mention a few. This paper reports on the problem of defining such a measure for precipitation and examines some practical methods for determining its statistical properties.

2. DEFINITION OF A SCALE MEASURE FOR PRECIPITATION EVENTS

We begin by considering the definition of precipitation events. In common weather service parlance, precipitation events are described in terms of depth of accumulation, duration, or both. For our purpose, we shall specify a precipitation event as an accumulation of water or water equivalent exceeding some threshold value during some fixed period of time. Both the threshold value and the period can be varied to fit any application.

We want to consider the size of the area covered by these events within some geographical region. An examination of a series of synoptic charts on which areas covered by precipitation events are delineated leaves one with an impression of the wildest sort of variability in the sizes and shapes of these patterns. At first glance, a rational characterization seems quite out of the question. However, we shall borrow a concept which is employed as a descriptor in other fields faced with the problem of describing essentially chaotic phenomena and define a characteristic size or scale measure of a precipitation event. The scale measure which we shall employ is that of the radius of the circle having an area equivalent to that covered by the phenomenon in question. See figure 1 for the geometric definition of the scale measure, r.

The most obvious and direct method of obtaining statistics on r would involve collecting a long series of daily precipitation maps, identifying the area covered by the event satisfying the stipulated criteria, measuring the area, computing r, and then tabulating the frequency function for values of r by geographical area, season, time of day, etc. A quick review of available data indicates that this approach would be quite impractical. The character of the spatial distribution of rainfall amounts would impose serious practical problems for the chartist, and the data required to carry out such a project are generally not available. For these reasons, it is useful to consider possibilities for an indirect method of obtaining information on the characteristics of the space scale of precipitation events.

We shall consider the problem of determining the relationship between the scale of some phenomenon P and the conditional probability that it does not occur at some preselected point, O, given that the phenomenon is detected somewhere within an area A equipped with a sampling network or a detection device. The geometry of this situation is depicted in figure 1.

In the diagram, R is the radius of the circle of observation whose area is A, r is the scale of the phenomenon P, the scale being defined as the radius of the circle of area equivalent to the area of P. If P is not circular in shape, then r is a general measure defining the average distance from the edge to the center of the area covered by P. If P is elliptical, r is the mean of the semimajor and semiminor axes. The domain of P is defined by the distribution of stations reporting a single event. We define the location of P as the location of the centroid,

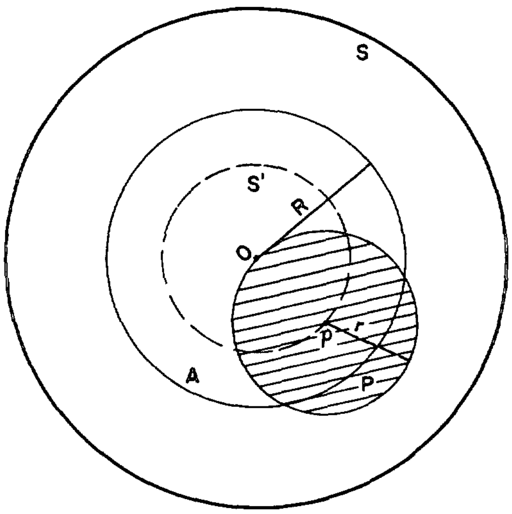

FIGURE 1.—Geometrical definitions of terms used in the problem of determining the space scale of precipitation events.

shown in the diagram as point p. S is the area in which p must lie for an occurrence of the phenomenon in A, and S' is the area excluded to p for a nonoccurrence at location O.

We define an occurrence of the phenomenon at any arbitrary location, O, as $O \in P$ and the nonoccurrence at O as $O \in P$, using set theory notation. For the situation depicted in the diagram, we shall determine first the probability of nonoccurrence at O given that P occurs somewhere in A. The latter condition being described in terms of the concept of intersection. The two criteria are $P \cap A \neq O$, $O \in P$. Now we make use of the conditional probability definition to write

$$\text{Prob}\,[O \in P | P \cap A \neq O] = \frac{\text{Prob}\,[p \in S, p \in S']}{\text{Prob}\,[p \in S]}.$$

We now assume that the probability of the location of point p falling within any area is proportional to the size of the area. Hence,

$$\text{Prob}\,[p \in S] = k\pi(R+r)^2,$$
$$\text{Prob}\,[p \in S, p \in S'] = k[\pi(R+r)^2 - \pi r^2],$$
$$\text{Prob}\,[O \in P | P \cap A \neq O] = \frac{(R+r)^2 - r^2}{(r+R)^2}, \text{ and}$$
$$1 - \text{Prob}\,[\quad] = \frac{r^2}{(R+r)^2} = Pr\,[\quad]. \tag{1}$$

Here $Pr[\ ]$ is used to denote the probability of the complementary event; that is, the occurrence of the phenomenon P at location O.

Measurements of the frequency of the complementary event are perhaps more common than those of the event itself, so we shall solve eq (1) in terms of the probability of the complementary event. For practical computation,

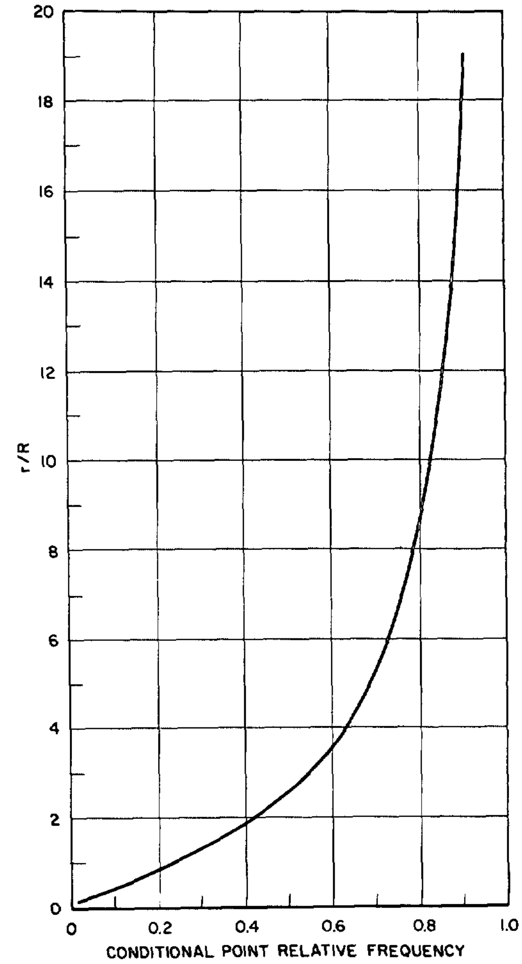

FIGURE 2.—The relationship between scale and conditional relative frequency.

we replace the probability by the observed relative frequency f and obtain from eq (1)

$$r/R = \frac{f \pm \sqrt{f}}{1-f}. \tag{2}$$

The solution of eq (2), using the positive root of f, is presented in graphical form in figure 2, the scale measure of P being expressed in terms of the ratio r/R.

Since variations in r arise from variations in both the size of the area covered by the event and its orientation, then, if the area is noncircular, r is a distributed element with some density function $h(r)$. The probability defined in eq (1) is therefore a stochastic variable which has a density function that can be expressed in terms of the parent density function, $h(r)$.

When r is distributed according to $h(r)$, Pr is distributed as $g(Pr)$, and the following relationship holds:

$$g(Pr) = h[r(Pr)] \left| \frac{dr}{dPr} \right|$$

where $r(Pr)$ is the inverse of eq (1). Solving eq (1) and substituting we obtain

$$g(Pr)=R\left[\frac{1+\frac{1}{2}(Pr)^{1/2}+\frac{1}{2}(Pr)^{-1/2}}{(1-Pr)^2}\right]h\left[\frac{R(Pr+(Pr)^{1/2})}{(1-Pr)}\right]. \tag{3}$$

We shall consider how information about $h(r)$ might be obtained from this approach.

We envision an observation program which will provide exact data on the number of occasions when a rain event occurred somewhere within the observing area and those also occurring at the central station. If we continue the assumption that the processes are random, then the expected number of such occurrences is obtained from the binomial probability function as

$$t=\sum_{i=1}^{j} n_i(Pr)_i$$

where n_i is the number of days on which a rain event of scale factor $r=r_i$ occurred in the area of observation, $(Pr)_i$ is the probability of rain at the central point as given by eq (1), and j is the total number of classes or values of scale measure used. Note that $\Sigma n_i=N$, the total number of occurrences sampled, and n_i/N is the value of the frequency function for $r=r_i$. If we denote this value as h_i, then

$$\hat{t}=h_1(Pr)_1+h_2(Pr)_2+h_3(Pr)_3+\ .\ .\ .\ ,$$

a relationship which shows how the frequency function for the scale factor operates on the point probability to establish the frequency of rain events at a location.

We now return to the problem of evaluating the frequency function for r using eq (3). If we assume the existence of statistics which provide the frequencies of precipitation events at a designated station and within an observational area over a long period of time, then values of $g(Pr)$ could be tabulated. Values of $g(Pr)$ and Pr can be introduced into eq (3) to obtain values for $h(r)$.

3. COMPUTATIONS FROM AVAILABLE DATA

To establish some characteristics of precipitation scale as we have defined it in eq (2), we used precipitation data for a subset of the National Weather Service cooperative precipitation network of the State of Kentucky (National Weather Records Center 1968). The data sample covers a 10-year period 1958–67 for the months of June, July, and August. The results are given in table 1 for an event defined as the occurrence of more than 0.09 in. of precipitation in the 24-hr period 0700–0700 LST. The network of stations used is shown in figure 3.

The scale measure depicted in the last three columns of table 1 is the radius of the circle of area equivalent to that of the precipitation event as we have defined it. The

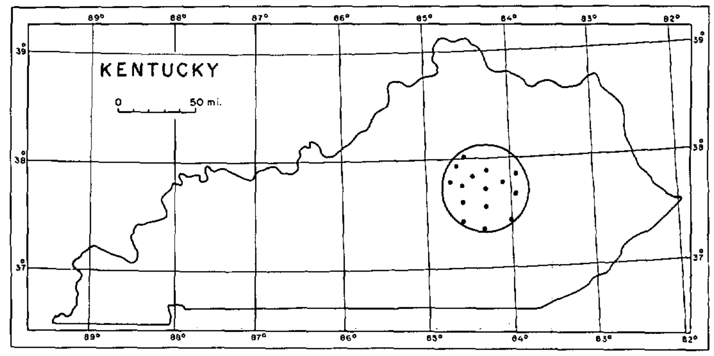

FIGURE 3.—Precipitation measurement network in Kentucky.

TABLE 1.—*Point, areal, and space-scale frequencies for summer rain events over a 2300 mi² area of Kentucky*

Date	Areal frequency of rain events			Conditional point frequency			Scale (mi)		
	June	July	August	June	July	August	June	July	August
1958	0.47	0.75	0.42	0.71	0.68	0.69	144.4	127.0	132.4
1959	.37	.69	.58	.55	.45	.34	77.5	55.0	37.8
1960	.50	.50	.52	.73	.67	.56	158.4	121.8	80.3
1961	.53	.67	.52	.56	.55	.56	80.3	77.5	80.3
1962	.50	.74	.29	.47	.52	.67	58.9	69.8	121.8
1963	.50	.57	.35	.67	.47	.46	121.8	58.9	56.9
1964	.30	.52	.42	.90	.50	.24	499.1	65.2	25.9
1965	.40	.55	.45	.33	.65	.43	36.4	112.3	51.4
1966	.23	.52	.51	.57	.44	.56	83.2	53.2	80.3
1967	.43	.66	.45	.69	.45	.50	132.4	55.0	65.2
MEAN	.43	.62	.44	.61	.55	.50	139.2	79.6	73.2
							*(96.5)	*(78.1)	*(73.0)

*These values are computed from the average of the frequencies

decrease in scale as the summer season advances is clearly shown, depicting the changing character of the precipitation regime. An important secondary property of the data in table 1 is the relatively high frequency of precipitation events recorded in the area during the month of July. This suggests an interesting feature of summertime precipitation: The threat of rain is higher (more precipitation cells are in existence) on any given day than in other seasons, but the point frequency is lower because of the decreasing characteristic scale of a rain event.

4. CONSIDERATIONS ON PRACTICAL APPLICATIONS

Previous sections of this paper have demonstrated how the scale measure for a precipitation event can be determined from properly organized climatological data from a network of precipitation recording stations. By varying the definition of an event (or phenomenon) in terms of either the period, amount, or both, much useful data can be generated on the characteristics of a precipitation regime for an area or season.

In the computations presented in this paper, a rather dense network of reporting stations was used. A less dense network can be used, but the use of data from widely separated stations raises the possibility that a significant number of events occurring over the sampling network might go undetected. For example, consider a network of k sampling stations within an area and let us determine the probability that a precipitation event will not be detected. The problem of determining this probability is very similar to the problem we have already solved.

The event space for nondetection of precipitation at any of the network stations, given that the event occurred somewhere in the area, is composed of the intersection of all events representing occurrences in the small areas surrounding each of the points but nonoccurrence at the network points. If we assume that the distribution of occurrences is random, the probability for k stations is, using notation previously defined,

$$(\text{Prob})_1 = (\text{Prob})^k. \tag{4}$$

This is the probability that a rain event will be completely missed by the sampling network.

The probability in eq (4), which can be used to correct the relative frequency in eq (2), is computed from eq (1), making use of the fact that each network station samples an area with radius approximately equal to $R/\sqrt{k}$. Hence,

$$(\text{Prob})_1 = \left[\frac{(R/\sqrt{k}+r)^2 - r^2}{(R/\sqrt{k}+r)^2}\right]^k \tag{5}$$

The best estimate for r/R is then obtained when a correction is applied to the value of f used in eq (2). The correction to f is obtained from eq (5) and should compensate to some extent for those rain events which pass undetected through the observing network.

Equation (2) is first solved with a value of f obtained from available data to establish an estimate of r. This value of r is used in eq (5) to estimate the frequency of missed rain events from $(\text{Prob})_1$. The best value of f to use for final estimates of r/R is $[1-(\text{Prob})_1]$ times the original value of f obtained from the records. Trial computations with eq (2) and (5) show that data from stations as widely spaced as those in the synoptic network are nearly useless for determining precipitation scale factors. For example, typical values like $R=150$ mi, $k=4$, Pr or $f=0.50$ yield a correction to f of only about 0.01.

In spite of the fact that we are unable to formulate a defensible rationale for iterating the correction process described above, repetitive applications of the procedure for a few cycles indicated neither convergence of the process nor significant changes in the scale measures obtained.

There is an interesting application of eq (2) to unconditional point frequency data. For example, the precipitation climatology for Lexington, Ky. indicates that the unconditional point frequency for the precipitation event for which frequencies are specified in table 1 is about 0.20 for the summer months. If we select our sampling radius to achieve an areal frequency of 1.0, we find, with eq (2), that R, the average distance to the centroid of a rain event, is $1.24r$. Since the distance to the edge of the event is $0.24r$, this indicates that in July, for example ($r=78$ mi), a precipitation event will occur, on the average, within 19 miles of any point in central Kentucky on any day in summer! This result seems incredible at first, but a moment's reflection on the data in table 1 provides a rather convincing argument for its reliability since the data show areal frequencies in excess of 0.05 for the 25-mi circle within which one would need to travel less than 25 mi to encounter a rain event. This appears to explain why during the dry part of summer, it frequently seems to rain everywhere except at the point where the observer is located. But the reduction in rain frequency results from a reduction in the scale of precipitation events.

5. RECOMMENDATIONS FOR FURTHER WORK

In view of the possible widespread uses which can be envisioned for a reliable scale measure for precipitation, further work should be directed toward defining the climatology of this element. Data on scale factors would add a new dimension to conventional descriptions of the precipitation regimes characterizing geographical areas and seasons, and provide additional guidance on questions relating to predictability.

REFERENCES

National Weather Records Center, ESSA, "Daily Precipitation Data for Kentucky Stations," Job 10512, Card Deck No. 486, Raleigh, N.C., June **1968** (raw data in files).

Roberts, Charles F., "Preliminary Results of an Empirical Study of Some Spectral Characteristics of Skill in Present Weather and Circulation Forecasts," *ESSA Technical Memorandum* WBTM FCST 14, Weather Analysis and Prediction Division, Silver Spring, Md., Nov. **1969**, 25 pp.

Roberts, Charles F., Porter, John M., and Cobb, Geraldine F., "Report on Weather Bureau Forecast Performance 1967–68 and Comparison With Previous Years," *ESSA Technical Memorandum* WBTM FCST 11, Weather Analysis and Prediction Division, Silver Spring, Md., Mar. **1969**, 44 pp.

[Received May 14, 1970; revised January 28, 1971]